

This is an electronic appendix to the paper by Allen 2003 The function of genomes in bioenergetic organelles. *Phil. Trans. R. Soc. Lond. B* **358**, 19-38.

Electronic appendices are refereed with the text. However, no attempt has been made to impose a uniform editorial style on the electronic appendices.

Electronic Appendix A: Protein-coding genes in chloroplast genomes.

Data of (Martin et al. 2002), as retrieved from the website <http://www.pnas.org/cgi/data/182432999/DC1/2>, with the minor modification that the genes have been sorted into the functional categories assigned by (Kaneko et al. 1996a; Kaneko et al. 1996b). Genes (rows) which did not have a clear homologue in cyanobacteria (that is, genes with a "-" in the column for Synechocystis) were not counted in the present analysis. For references to notes and comments, see <http://www.pnas.org/cgi/data/182432999/DC1/2>. For direct links to the corresponding genomes and individual amino acid sequences, see [http://www.molevol.de/people/martin/projects/how\\_many/Plastid\\_Gene\\_Table/AA\\_table.html](http://www.molevol.de/people/martin/projects/how_many/Plastid_Gene_Table/AA_table.html).

**Abbreviations and Genome Accession Numbers:** **Zea**, *Zea mays* (X86563); **Ory**, *Oryza sativa* (X15901); **Nic**, *Nicotiana tabacum* (S54304); **Spi**, *Spinacia oleracea* (AJ400848); **Oen**, *Oenothera elata* (NC\_002693); **Pin**, *Pinus thunbergii* (D17510); **Mar**, *Marchantia polymorpha* (X04465); **Eug**, *Euglena gracilis* (NC\_001603); **Chl**, *Chlorella vulgaris* (NC\_001865); **Nep**, *Nephroselmis olivacea* (NC\_000927); **Mes**, *Mesostigma viride* (NC\_002186); **Od**, *Odontella sinensis* (Z67753); **Gui**, *Guillardia theta* (NC\_000926); **Por**, *Porphyra purpurea* (U38804); **Can**, *Cyanidium caldarium* (AF022186, Z36235, Z70297); **Cpa**, *Cyanophora paradoxa* (NC\_001675); **Syn**, *Synechocystis PCC 6803* (NC\_000911); sim., similar to; hom., homologue; SU, subunit; bind., binding; prot., protein; RT, reverse transcriptase.

	<b>Z</b>	<b>O</b>	<b>N</b>	<b>S</b>	<b>O</b>	<b>P</b>	<b>M</b>	<b>E</b>	<b>C</b>	<b>N</b>	<b>M</b>	<b>O</b>	<b>G</b>	<b>P</b>	<b>C</b>	<b>C</b>	<b>S</b>	nuclear homologue	Comments
	<b>e</b>	<b>r</b>	<b>i</b>	<b>p</b>	<b>e</b>	<b>i</b>	<b>a</b>	<b>u</b>	<b>h</b>	<b>e</b>	<b>d</b>	<b>u</b>	<b>o</b>	<b>a</b>	<b>y</b>	<b>y</b>	<b>a</b>	<b>n</b>	
<b>Amino acid biosynthesis</b>																			
<i>argB</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	pir T46192	N-Acetylglutamate kinase	
<i>carA</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AY046004	Carbamoyl phosphate synthetase small SU	
<i>cysA</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Probable transport prot. ( <i>mbpX</i> )	
<i>cysT</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Probable transport prot. ( <i>mbpY</i> )	
<i>gltB</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AL396716	Glutamate synthase (GOGAT)	
<i>hisH</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	sp Q9SZ30	Histidinol-phosphate aminotransferase; <i>Syn</i> slr0084	
<i>ihvB</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	sp P09114	Acetohydroxyacid synthase large SU	
<i>ihvH</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AB005242	Acetohydroxyacid synthase small SU	
<i>trpA</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AL139656	Tryptophane synthase $\alpha$ SU	
<i>trpG</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AC004133	Anthraniolate synthase component II, glutamine amidotransferase	

## Biosynthesis of cofactors and prosthetic groups





<i>petM</i>	sp Q42496	<i>ycf3I</i> ; Cytochrome <i>b6f</i> /complex SU VII <i>ycf6</i> ; Syn sml0004 (see Hager et al. 1999, EMBO J. 18:5834)
<i>petN</i>		
<i>psaA</i>		PSI P700 apoprotein A1
<i>psaB</i>		PSI P700 apoprotein A2
<i>psaC</i>		PSI iron-sulphur center FA/FB containing SU VII
<i>psaD</i>		sp P36213 PSI ferredoxin-bind. prot. SU II
<i>psaE</i>		sp P13194 PSI SU IV, 18-20 kDa
<i>psaF</i>		sp P12355 PSI SU III, plastocyanin-bind. prot.
<i>psaI</i>		PSI SU VIII
<i>psaJ</i>		PSI SU IX
<i>psaK</i>		PSI SU X
<i>psaL</i>		PSI SU XI, reaction center prot.
<i>psaM</i>		PSI M-polypeptide
<i>psbA</i>		PSII D1 reaction center prot.
<i>psbB</i>		PSII CP47 chlorophyll apoprotein
<i>psbC</i>		PSII CP43 chlorophyll apoprotein
<i>psbD</i>		PSII D2 reaction center prot.
<i>psbE</i>		PSII cytochrome <i>b559</i> α SU
<i>psbF</i>		PSII cytochrome <i>b559</i> β SU
<i>psbH</i>		PSII 10 kDa phosphoprotein
<i>psbI</i>		PSII I polypeptide
<i>psbJ</i>		PSII J protein
<i>psbK</i>		PSII K protein
<i>psbL</i>		PSII L protein
<i>psbM</i>		PSII M protein
<i>psbN</i>		PSII N protein
<i>psbS</i>	sp Q02060	<i>ycf17</i> , Syn sll1633, ELIP superfamily (Grimm et al., 1989), spinach PS II 22 kD protein
<i>psbT</i>		<i>ycf8</i> , PSII T protein
<i>psbV</i>		PSII cytochrome c550 ( <i>petK</i> ) <i>psbW</i> Gene nomenclature problem: see <i>ycf79</i> and footnote
		<i>ycf79</i> , 13 kDa PS II protein, <i>Syn</i> sll1398, see footnote
		<i>ycf32</i> , PSII core complex prot. spinach, <i>Syn</i> sml0007
		<i>Syn</i> sml0002 PSII X 4.1 kDa protein
		<i>ycf9</i> , Syn sll1281, see Swiatek et al. Plant Cell 13:1347-1367 ( <i>ycf9</i> in Can is highly divergent but scored "+")
		Rubisco large SU, green type
		Rubisco large SU, red type
		<i>ycf30</i> ; Syn sll0998 LysR family of transcriptional regulators



<i>rpl11</i>	AB042934	Ribosomal protein L11
<i>rpl12</i>	sp P02398	Ribosomal protein L12
<i>rpl13</i>	sp P12629	Ribosomal protein L13
<i>rpl14</i>		Ribosomal protein L14
<i>rpl16</i>		Ribosomal protein L16
<i>rpl18</i>		Ribosomal protein L18
<i>rpl19</i>	AF336922	Ribosomal protein L19
<i>rpl20</i>	AF250384	Ribosomal protein L20
<i>rpl21</i>	sp P51412	Ribosomal protein L21
<i>rpl22</i>		Ribosomal protein L22
<i>rpl23</i>	sp P11893	Ribosomal protein L23
<i>rpl24</i>	sp P30155	Ribosomal protein L24
<i>rpl27</i>	P30596	Ribosomal protein L27
<i>rpl28</i>	AF147725	Ribosomal protein L28
<i>rpl29</i>	AC023754	Ribosomal protein L29
<i>rpl31</i>		Ribosomal protein L31
<i>rpl32</i>		Ribosomal protein L32
<i>rpl33</i>	sp P233326	Ribosomal protein L33
<i>rpl34</i>		Ribosomal protein L34
<i>rpl35</i>	X93156	Ribosomal protein L35
<i>rpl36</i>	AF250383	Ribosomal protein S5
<i>rps1</i>	P29344	Ribosomal protein S1
<i>rps2</i>		Ribosomal protein S2
<i>rps3</i>		Ribosomal protein S3
<i>rps4</i>		Ribosomal protein S4
<i>rps5</i>		Ribosomal protein S5
<i>rps6</i>		Ribosomal protein S6
<i>rps7</i>		Ribosomal protein S7
<i>rps8</i>		Ribosomal protein S8
<i>rps9</i>	AB022676	Ribosomal protein S9
<i>rps10</i>	AP000375	Ribosomal protein S10
<i>rps11</i>		Ribosomal protein S11
<i>rps12</i>		Ribosomal protein S12
<i>rps13</i>	sp P42732	Ribosomal protein S13
<i>rps14</i>		Ribosomal protein S14
<i>rps15</i>		Ribosomal protein S15
<i>rps16</i>	sp P17092	Ribosomal protein S16
<i>rps17</i>		Ribosomal protein S17
<i>rps18</i>		Ribosomal protein S18
<i>rps19</i>		Ribosomal protein S19



## REFERENCES

- Allen, J. F. 2003 The function of genomes in bioenergetic organelles. *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.*, in press.  
 Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hiroswa, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. & Tabata, S. 1996a Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109-36.

- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sasaki, M., Sugiura, M., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpoo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. & Tabata, S. 1996b Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions (supplement). *DNA Res.* **3**, 185-209.
- Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M. & Penny, D. 2002 Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12246-12251.